

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 19:43:59 ; Search time 9.25714 Seconds
(without alignments)
2096.859 Million cell updates/sec

Title: US-09-497-967-7

Perfect score: 2540

Sequence: 1 MKNILVILLISFINQIKS.....QCDFANFLSISLLISYLL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	236.5	9.3	667	TS11_GIALA	Q03185 giardia lam
2	228.5	9.0	713	TSAA_GIALA	P21849 giardia lam
3	189	7.4	328	CI70_GIALA	P15799 giardia lam
4	188	7.4	1609	LMG1_HUMAN	P11047 homo sapien
5	188	7.4	5376	ZAN_MOUSE	O88799 mus musculus
6	185.5	7.3	1696	PK5_MOUSE	O98115 branchiost
7	184.5	7.3	687	VS41_GIALA	P92127 giardia lam
8	183.5	7.2	1680	FUR2_DROME	P30432 drosophila
9	179.5	7.1	1877	PK5_MOUSE	Q04592 mus musculus
10	179	7.0	2715	G156_PARPR	P13837 paramecium
11	178.5	7.0	1607	LMG1_MOUSE	P02468 mus musculus
12	175.5	6.9	3075	LMAL_HUMAN	P25391 homo sapien
13	174.5	6.9	3718	LMA_MOUSE	Q61001 mus musculus
14	173	6.8	3712	LMA_DROME	Q00174 drosophila
15	167.5	6.6	2704	G168_PARPR	P17053 paramecium
16	166.5	6.6	969	PAC4_HUMAN	P29122 homo sapien
17	166	6.5	3084	LMAL_MOUSE	P19137 mus musculus
18	165	6.5	1700	BAR3_CHITE	Q03376 chironomus
19	162	6.4	1576	YLK3_CAEEL	P41951 caenorhabdi
20	158	6.2	3106	LMAL_MOUSE	Q60675 mus musculus
21	157	6.2	3110	LMAL_HUMAN	P24043 homo sapien
22	155.5	6.1	1246	YVY2_CAEEL	P34504 caenorhabdi
23	155.5	6.1	3672	LM2_CAEEL	Q21313 caenorhabdi
24	154	6.1	1581	LMG3_MOUSE	Q91066 mus musculus
25	153.5	6.0	1790	LMB1_DROME	P11046 drosophila
26	150.5	5.9	670	VG50_HSVII	Q00130 ictaluriid h
27	150.5	5.9	1169	YK82_YEAST	P36170 saccharomyc
28	150	5.9	1168	LMB3_MOUSE	Q61087 mus musculus
29	148.5	5.8	1639	LMG1_DROME	P15215 drosophila
30	148	5.8	3695	LMA5_HUMAN	O15230 homo sapien
31	147	5.8	712	FBL1_CAEEL	O77469 caenorhabdi
32	146.5	5.8	937	PAC4_MOUSE	Q63415 rattus norv
33	146.5	5.8	1193	LMG2_HUMAN	Q13753 homo sapien

34	145.5	5.7	600	1	SP96_DICDI	P14328 dictyostell
35	145.5	5.7	790	1	ANP_NOTCO	P24856 notothenia
36	143.5	5.6	2471	1	NTC2_RAT	O9q30 rattus norv
37	142	5.6	1167	1	XMRK_XIPMA	P13388 xiphophorus
38	142	5.6	1587	1	LMG3_HUMAN	Q9y66 homo sapien
39	141.5	5.6	610	1	LEM2_HUMAN	P16581 homo sapien
40	141.5	5.6	1786	1	LMB1_MOUSE	P02469 mus musculus
41	141	5.6	2911	1	FBN2_HUMAN	P35556 homo sapien
42	140.5	5.5	1416	1	YN81_CAEEL	Q03610 caenorhabdi
43	140	5.5	1064	1	FBP1_SRPUP	P10079 strongyloce
44	139.5	5.5	677	1	SP87_DICDI	P54643 dictyostell
45	139	5.5	768	1	LEM3_MOUSE	Q01102 mus musculu

ALIGNMENTS

RESULT 1
TS11_GIALA STANDARD: PRT; 667 AA.
AC Q03185;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Major surface trophozoite antigen II precursor.
GN TSP11.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate AD-1;
RX MEDLINE=932411215; PubMed=8479449;
RA Ey P.L., Khanna K., Manning P.A., Mayrhofer G.;
RT "A gene encoding a 69-kilodalton major surface protein of Giardia intestinalis trophozoites.";
RL Mol. Biochem. Parasitol. 58:247-258(1993).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE PLASMA MEMBRANE.
CC -!- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC EMBL; M95814; AAA02687.1; -.
DR PIR; A48579; A48579.
DR HSP; P02468; IKLO.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00261; FU; 5.
DR Antigen; Repeat; Transmembrane; Signal.
FT SIGNAL 1 17
FT CHAIN 18 667 MAJOR SURFACE TROPHOZOITE ANTIGEN 11.
FT DOMAIN 18 633 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 634 662 POTENTIAL.
FT DOMAIN 663 667 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 667 AA; 68475 MW; 1DD9572703232B8D CRC64;

Query Match 9.3%; Score 236.5; DB 1; Length 667;
Best Local Similarity 22.7%; Pred. No. 7.5e-10;
Matches 128; Conservative 49; Mismatches 193; Indels 195; Gaps 30;

[illegible]

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M55210; AAA59492.1; JOINED.
DR EMBL; M55217; AAA59492.1; JOINED.
DR EMBL; M55201; AAA59492.1; JOINED.
DR EMBL; M55211; AAA59492.1; JOINED.
DR EMBL; M55212; AAA59492.1; JOINED.
DR EMBL; M55213; AAA59492.1; JOINED.
DR EMBL; M55214; AAA59492.1; JOINED.
DR EMBL; M55215; AAA59492.1; JOINED.
DR EMBL; M55216; AAA59492.1; JOINED.
DR EMBL; M55192; AAA59492.1; JOINED.
DR EMBL; M55193; AAA59492.1; JOINED.
DR EMBL; M55194; AAA59492.1; JOINED.
DR EMBL; M55195; AAA59492.1; JOINED.
DR EMBL; M55196; AAA59492.1; JOINED.
DR EMBL; M55197; AAA59492.1; JOINED.
DR EMBL; M55198; AAA59492.1; JOINED.
DR EMBL; M55199; AAA59492.1; JOINED.
DR EMBL; M55200; AAA59492.1; JOINED.
DR EMBL; M55202; AAA59492.1; JOINED.
DR EMBL; M55204; AAA59492.1; JOINED.
DR EMBL; M55205; AAA59492.1; JOINED.
DR EMBL; M55206; AAA59492.1; JOINED.
DR EMBL; M55207; AAA59492.1; JOINED.
DR EMBL; M55208; AAA59492.1; JOINED.
DR EMBL; M55209; AAA59492.1; JOINED.
DR EMBL; J03202; AAA59488.1; -.
DR EMBL; M27654; AAA59489.1; -.
DR EMBL; X13939; CAA32122.1; -.
DR PTR; S13548; MMH092.
DR HSSP; P02468; 1TLE.
DR Genew; HGNC:6492; LAMC1.
DR MIM; 150290; -.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00052; laminin_B_1.
DR Pfam; PF00053; laminin_EGF; 10.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 1.
DR SMART; SM00180; EGF Lam; 10.
DR SMART; SM00281; LamB; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR Glycoprotein; Basemem membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 33
FT CHAIN 34 1609 LAMININ GAMMA-1 CHAIN.
FT DOMAIN 34 285 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 286 341 LAMININ EGF-LIKE 1.
FT DOMAIN 342 397 LAMININ EGF-LIKE 2.
FT DOMAIN 398 444 LAMININ EGF-LIKE 3.
FT DOMAIN 445 494 LAMININ EGF-LIKE 4.
FT DOMAIN 495 504 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 505 689 LAMININ DOMAIN IV.
FT DOMAIN 690 723 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT DOMAIN 724 772 LAMININ EGF-LIKE 6.
FT DOMAIN 773 827 LAMININ EGF-LIKE 7.
FT DOMAIN 828 883 LAMININ EGF-LIKE 8.
FT DOMAIN 884 934 LAMININ EGF-LIKE 9.
FT DOMAIN 935 982 LAMININ EGF-LIKE 10.
FT DOMAIN 983 1030 LAMININ EGF-LIKE 11.
FT DOMAIN 1030 1609 DOMAIN II AND I.
FT DOMAIN 1038 1609 COILED COIL (POTENTIAL).

FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 288 305 BY SIMILARITY.
FT DISULFID 307 316 BY SIMILARITY.
FT DISULFID 319 339 BY SIMILARITY.
FT DISULFID 342 351 BY SIMILARITY.
FT DISULFID 344 367 BY SIMILARITY.
FT DISULFID 370 379 BY SIMILARITY.
FT DISULFID 382 395 BY SIMILARITY.
FT DISULFID 398 410 BY SIMILARITY.
FT DISULFID 400 416 BY SIMILARITY.
FT DISULFID 418 427 BY SIMILARITY.
FT DISULFID 430 442 BY SIMILARITY.
FT DISULFID 445 456 BY SIMILARITY.
FT DISULFID 447 463 BY SIMILARITY.
FT DISULFID 465 474 BY SIMILARITY.
FT DISULFID 477 492 BY SIMILARITY.
FT DISULFID 724 733 BY SIMILARITY.
FT DISULFID 726 740 BY SIMILARITY.
FT DISULFID 742 751 BY SIMILARITY.
FT DISULFID 754 770 BY SIMILARITY.
FT DISULFID 773 781 BY SIMILARITY.
FT DISULFID 775 792 BY SIMILARITY.
FT DISULFID 795 804 BY SIMILARITY.
FT DISULFID 807 825 BY SIMILARITY.
FT DISULFID 828 842 BY SIMILARITY.
FT DISULFID 830 849 BY SIMILARITY.
FT DISULFID 852 861 BY SIMILARITY.
FT DISULFID 864 881 BY SIMILARITY.
FT DISULFID 884 898 BY SIMILARITY.
FT DISULFID 886 905 BY SIMILARITY.
FT DISULFID 907 916 BY SIMILARITY.
FT DISULFID 919 932 BY SIMILARITY.
FT DISULFID 935 947 BY SIMILARITY.
FT DISULFID 937 954 BY SIMILARITY.
FT DISULFID 956 965 BY SIMILARITY.
FT DISULFID 968 980 BY SIMILARITY.
FT DISULFID 983 995 BY SIMILARITY.
FT DISULFID 985 1001 BY SIMILARITY.
FT DISULFID 1003 1012 BY SIMILARITY.
FT DISULFID 1015 1028 BY SIMILARITY.
FT DISULFID 1031 1031 INTERCHAIN (PROBABLE).
FT DISULFID 1034 1034 INTERCHAIN (PROBABLE).
FT DISULFID 1600 1600 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 60 60 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 650 650 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1175 1175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1205 1205 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1223 1223 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1241 1241 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1380 1380 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1395 1395 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 212 212 F -> I (IN REF. 2).
SQ SEQUENCE 1609 AA; 177606 MW; B098F20FCD97293B CRC64;

Query Match 7.4%; Score 188; DB 1; Length 1609;
Best Local Similarity 23.9%; Pred. No. 5.4e-06;
Matches 99; Conservative 28; Mismatches 165; Indels 122; Gaps 26;

QY 75 AGAQNPPATAMLVQCNVKCPAGTAIAGGATDYAAIITECVNCRINFENENAPNFNAGA 134
Db 674 ASARPGGVPATWVESCT--CPVG---YGGQF-----CEMC-LSGYRRETPNLGP-Y 718
QY 135 STCTACPNVRVGGALTAGNAATIVAQC--NVACPTGTALDDGVTTDYVRSFTECVKCRIN 192
Db 719 SPCVLCAKN---GHSTCDPFGVCNCRNTAGP-----HCEKCSDG 757
QY 193 FYYNGNNGNTPFNPGKSQCTPCP-----AIKP-----ANVAQATLGNDATITAOQN 238


```
FT CARBOHYD 1908      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1933      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2028      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2111      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2142      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2332      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2533      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2575      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2692      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2812      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3052      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3065      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3144      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3172      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3288      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3292      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3782      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4005      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4136      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4243      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4254      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4335      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4376      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4586      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 5136      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 5252      N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 5376 AA; 579908 MW; 0E44DB77DF2A2620 CRC64;

Query Match      7.4%; Score 188; DB 1; Length 5376;
Best Local Similarity 21.2%; Pred. No. 1.8e-05;
Matches 114; Conservative 52; Mismatches 184; Indels 188; Gaps 30;

QY 17 QLSANCPVGTETNAGQVDDLGTTPANCVN---CQKNFYNNAAAFVPGASTCTP-C-- 69
Db 3151 QCNQCQCPKTYCK-----DLKSSNCTNIPLOCPAHSRTNC-----LPSPCPCLD 3199

QY 70 POKKDAGAQPNPATANLVTCQNV-----KCPAGTAIAGGATDYAAIITECVNCRIWF 122
Db 3200 PEGLCGTSKVPSTCEGCIQCPGLMHKNC-----VLRIFGCKNTQCAF 3247

QY 123 YNENAPFNAGASTCTACPNRVGGALTAGNAATIVAQC-NVACPTGTALDDGVTTDYVR 181
Db 3248 ISADKTWISRGCTQSTCTP---AGAI-----HCRNFKCPSGT----- 3281

QY 182 SFTECVKRLNFYNGNGNTPFN-----PKSQCTPC-----PAIKP 219
Db 3282 -----YCKNGDNGSSNCTEITLQCPINSQFTDCLPSCVPSCSNRCVETSPSV-P 3329

QY 220 ANVAQATLGNDAFI-----TAQCNVACPDGTISAAGVNNVAAQ-----NTE 260
Db 3330 SSREGCLCNHGTFVSEDKCVPRTCQCKDARGAIIIPAG-KTWTSKGCTQSCACVEGNIQ 3388

QY 261 CTN--CAPNFYNNAPNPNFGNSTC---LPCPANKDYGAETA-----GGAATLA 305
Db 3389 CNFQCPPEY---CKDNSEGSSTKITLQCPAHTQYTSCLPCLPCLDPEGLCKDKDIS 3445

QY 306 KCQNIACPDGTAIAGATNVYIQLQTECLNCAANFYFDGNNFQAQSSRCKACPAKRVQAV 365
Db 3446 PKVPSTCKEGCVQSG---YVLNSDKCVLRA-----ECDCRDAQAL 3484

QY 366 ATAGGTATL-----TAQC-ALECPAGTVLTDTSTYKQAASECVK-----CAA 408
Db 3485 IPAGKTWTSPTGCTQSCACMGVAGQOSSQCPGTYCKDN-----EDGNSNCAKITLQCPA 3539

QY 409 NFYTTKQTDWAGIDTC-TSC--NKKLTSGABANLPESAKKNICQDFANFLSISILLI 463
Db 3540 HSLFTN-----CLPPCLPCLDPLDGLCKGASPKVPSTCKEGCICQSGYVLSNNCKLL 3591

RESULT 6
PCK5_BRACL
ID PCK5_BRACL STANDARD; PRT; 1696 AA.
AC Q9NJ15; Q9NJ16; Q9NJ14;
```

```
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Proprotein convertase PC6-like) (apc6).
GN PC6.
OS Branchiostoma californiensis (Callinomia lancelet) (Amphioxus).
OS Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7738;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RP MEDLINE=20175281; PubMed=10708868;
RX Oliva A.A. Jr., Chan S.J., Steiner D.F.;
RT "Evolution of the prohormone convertases: identification of a
RL homologue of pc6 in the protochordate amphioxus.";
CC Blochim. Biophys. Acta 1477:338-348(2000).
CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS
CC -!- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF184615; AAF26300.1; -
CC EMBL; AF184616; AAF26301.1; -
CC EMBL; AF184617; AAF26302.1; -
CC HSSP; Q99405; 1MPT.
CC MEROPS; S08.UPB; -.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR002884; P_domain.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC Pfam; PF01483; P; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PRODOM; PD000717; P_domain; 1.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00261; FU; 17.
CC PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Cleavage on pair of basic residues; Repeat; Alternative splicing;
KW Transmembrane.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 110 POTENTIAL.
FT CHAIN 111 1696 PROPROTEIN CONVERTASE SUBTILISIN/KEXIN
FT TYPE 5.
FT DOMAIN 111 1618 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1619 1639 POTENTIAL.
FT DOMAIN 1640 1696 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 111 488 CATALYTIC.
FT DOMAIN 496 637 HOMO B.
FT DOMAIN 664 1649 CYS-RICH MOTIF (CRM) REGION.
FT SITE 110 111 CLEAVAGE (AUTO-) (BY SIMILARITY).
```

```
FT ACT_SITE 192 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1259 1323 DFTLDRGCLITSCGPGGYMDREKKRACHTCKECSDEY
FT DFTCTACNCGFLITDASSCAGCP -> AENQNASFCFFA
FT PREVSIAELALHGLRSLTDPQSNPPDVLGADRRL
FT TTATSAAGRA (IN ISOFORM C).
FT VARSPLIC 1324 1696 CHPTCKECSDEYDCTACNCGFLITDASSCAGCPQOFL
FT VARSPLIC 1288 1343 HGGDCDSCHREKTC -> IARCVDRRBSWCDLVRNFC
FT VRYFVKRCCTCKLYMDRPNRRSSQFTQGRN (IN
FT ISOFORM A).
FT VARSPLIC 1344 1696 MISSING (IN ISOFORM A).
FT SEQUENCE 1696 AA; 188410 MW; 281CBEL1784257CBD CRC64;

Query Match 7.3%; Score 185.5; DB 1; Length 1696;
Best Local Similarity 20.4%; Pred. No. 8.6e-06;
Matches 101; Conservative 43; Mismatches 189; Indels 161; Gaps 25;

QY 18 IKSANCPVGTETAG-----QVDDLGPANVCNQKNFYNNAAAFVPGASTCT 67
DB 1048 VRTNCPSFTYDQDRECPCHDNCACDGNPNQNCCKEGFYKT-----PDGCS--T 1100
QY 68 PCPO---KKDAGAGPNP-----PATANLV-----TQCNVYKCPAGTAIAGG 104
DB 1101 GCPNRYKDDTNECKPCDSSCFCTSGPASFCLSACDGFLEHSSCRSTCPAG--FTGN 1158
QY 105 ATDYAALITCVNCRINFYNENAPNPNAGASTCTACPVNRVGGALTAGNAATIVAQCINVA 164
DB 1159 AESHECVESCC-----EQDYYSSETGRCEDCPN-----CRACNDGDCAC 1201
QY 165 CPTGTALD-----DGVTTDYVRSFTECVKRLNFYNNNGNTFFNPGKSQCTPC 214
DB 1202 APIYIVVDGRCEPEETCEGEGVQDRDRDAE--LSC-----PCHQSKCTC 1245
QY 215 PAKPANVAQATLGNATITACNVACPDGTISAAGVNNVNAQNTPEC--TNCAPNFY--NNN 272
DB 1246 SG--PSD-----TDCDSCKGDDTILDRG-----ECITSCGPGFYMDRR 1281
QY 273 APNENPGNSTCLPCPANKDYGAEATA---GGAATLAKGNTACPDGTGTAIAAGATNYVLQ 329
DB 1282 EKCKKACHTCKEC--SDEYDTCACNDGFLITDASSCEAGCPGPGFLHHG----- 1331
QY 330 TECLNCAANFYFDGNFQAGSSRCRACAPANKVQGAATAGGTATLTAQCALECPAGTVLT 389
DB 1332 -DCDSC-----HRECKTCGPHDNCNLSQPGSYLNDQOCSTHCPEGTF-- 1374
QY 390 DGTSTYKQAAEF---CVKCAANFYTKQTDWVAGIDTCTSC-----NKKLTSGA 436
DB 1375 ---EETVDSGETVLQRLCHVNCCKTCHG---EGEEDCEACANDIKYKQDGRCVTECQ 1427
QY 437 EANLPESAKKNIQC 450
DB 1428 EGHYPLTNECQC 1441

RESULT 7
VS41_GIALA
ID VS41_GIALA STANDARD; PRT; 687 AA.
AC P92127;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Variant-specific surface protein VSP4A1 precursor (CRISP-90).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O2-4A1;
```

```
RX MEDLINE-97321554; PubMed-9178264;
RA Papanastasiou P., Bruderer T., Li Y., Bomelli C., Koehler P.;
RT "Primary structure and biochemical properties of a variant-specific
RT surface protein of Giardia.";
RT Mol. Biochem. Parasitol. 86:13-27(1997).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE-97233006; PubMed-9078242;
RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
RT "The variant-specific surface protein of Giardia, VSP4A1, is a
RT glycosylated and palmitoylated protein.";
RT Biochem. J. 322:49-56(1997).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -!- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
CC AT THE REDUCING TERMINUS.
CC -!- PTM: PALMITOYLATED.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z83743; CAB06038.1; -.
CC HSSP; O14763; ID0G.
CC GLYCOSuiteDB; P92127; -.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR005127; Giardia_VSP.
CC Pfam; PF03302; VSP; 2.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00001; EGF-like; 1.
CC SMART; SM00261; FU; 3.
CC
CC Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
CC Signal.
CC SIGNAL 1 14 POTENTIAL.
CC CHAIN 15 687 VARIANT-SPECIFIC SURFACE PROTEIN VSP4A1.
CC DOMAIN 15 660 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 661 681 POTENTIAL.
CC DOMAIN 682 687 CYTOPLASMIC (POTENTIAL).
CC SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;
CC
Query Match 7.3%; Score 184.5; DB 1; Length 687;
Best Local Similarity 21.1%; Pred. No. 4.1e-06;
Matches 119; Conservative 54; Mismatches 214; Indels 177; Gaps 29;

QY 12 SLFINQIKSANCPVGTET---NTAGQVDDLTGTPANVCNOKNFYNNAAAFVPGASTCT 67
DB 124 SQIFQNKATPSEKSEICILCWDITDRNGVGV-ANCATC-----TAPASSTGPATCT 175
QY 68 PCP---OKKDAGAOPNPATANLVTCNVKCPAGTAGGATDYAALITCVNCRINFY 123
DB 176 ECMAGTYKKSD-----TEC-AACHSDCATCSGEAN-----NOCTSCETGKY 215
QY 124 -----NENAPNFN---AGASTCTACPVNRVGGALTAGNAATIVAQCNVACPTG--- 168
DB 216 LKSNQCVKEKNTCNTNHYDDTSMTCVACTVLDANCATCSFDSATAKGC-LTCNSNKIPR 274
QY 169 TALDDGVTTDYVRSFTECVKRLNFYNNNGNTFFNPGKSQCTPCPAIKPANVAQNTLG 228
DB 275 TTL-DGTSTCVENSYAGC-----QGADNELFMKEDQSACLLCGDTEKNSDKG-VA 323
QY 229 NDATITACQN-----VACPDGTISAAG-VNNVNAQN-----TECTNCAPNFYNN 271
DB 324 NCRTCTKNANDSPPTCTACLDGYFLERGSCTTTCCADNCATCSEATTEKCKICKAGFF-- 381
QY 272 NAPNPNPCNSTCLPCPANKDYG----AEATAGGAATL-AKQC-----NIACPDG 315
DB 382 ---LASPGEGKICISDNTNNGIDGCAECTKEPAGPLKCTCKPNRKPACTSDNYTCTEK 438
```

```
Qy 316 TAI-----ASGATNVILOTE-----C 332
| ||| : ||
Db 439 TCDEPTVCGGTSGACDAIVDANGKEHYCSYCGETNKEFPIDGLCTDNKGTNAGCTDHTC 498
| ||| : ||
Qy 333 LNCANFYFGNFFQAGSSRCCKACPAKQVQAVATAGGTATLIA-----QC 378
| ||| : ||
Db 499 SYCAAGFEL-----YMGCKYKIDVPGSYMCKSTTAGVCDTPNANNRFFVVPKALSAQS 554
| ||| : ||
Qy 379 ALEC--PAGTVLTDGTTSTYKQAASECVCKAANFYTTKQTDWVAGIDTCTSCNKKLTSGA 436
| ||| : ||
Db 555 VLACGNPLGII-AGNAYVVEGSCQCTAPDA-----RADGGMVATCTACEDGKPG- 606
| ||| : ||
Qy 437 EANLPESAKNIOCFANFISLSL 460
| ||| : ||
Db 607 -----KSGTGCVACPDANCKSCTM 625

RESULT 8
ID FUR2_DROME STANDARD; PRT; 1680 AA.
AC P30432;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Furin-like protease 2 precursor (EC 3.4.21.75) (Furin 2).
GN FUR2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92381036; PubMed=1512259;
RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,
RA Rentrop M., Gateff E.A.F., Leunissen J.A.M., van de Ven W.J.;
RT "Cloning and functional expression of Dfurin2, a subtilisin-like
RT proprotein processing enzyme of Drosophila melanogaster with multiple
RT repeats of a cysteine motif."
RT J. Biol. Chem. 267:17208-17215(1992).
CC -!- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPEPTASE
CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF
CC CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their
CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
CC complement component C3 and von Willebrand factor from their
CC respective precursors.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94375; AAA28551.1; -.
CC F01434; A43434.
CC HSP; Q99405; IMPT.
CC MEROFS; S08.049; -.
CC FlyBase; FBgn0004598; Fur2.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR002884; P_domain.
CC InterPro; IPR002809; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC Pfam; PF01483; P; PARTIAL.
CC PRINTS; PR00723; SUBTILISIN.
CC ProDom; PD000717; P_domain; 1.
CC SMART; SM00181; EGF; 1.
```

```
DR SMART; SM00261; FU; 10.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal; Transmembrane;
KW Multigene family; Zymogen; Repeat.
FT SIGNAL 1 ?
FT PROPEP 319
FT CHAIN 320 1680
FT ACT_SITE 418 418
FT ACT_SITE 457 457
FT ACT_SITE 638 638
FT DOMAIN 962 1444
FT REPEAT 962 1007
FT REPEAT 1008 1057
FT REPEAT 1058 1104
FT REPEAT 1105 1153
FT REPEAT 1154 1205
FT REPEAT 1206 1254
FT REPEAT 1255 1299
FT REPEAT 1300 1346
FT REPEAT 1347 1393
FT REPEAT 1394 1444
FT TRANSMEM 1508 1532
FT DOMAIN 1533 1680
FT CARBOHYD 3 3
FT CARBOHYD 109 109
FT CARBOHYD 130 130
FT CARBOHYD 203 203
FT CARBOHYD 443 443
FT CARBOHYD 481 481
FT CARBOHYD 928 928
FT CARBOHYD 1061 1061
FT CARBOHYD 1182 1182
FT CARBOHYD 1275 1275
FT CARBOHYD 1278 1278
FT CARBOHYD 1440 1440
FT SEQUENCE 1680 AA; 183599 MW; 0A99CE8770A8E293 CRC64;
Query Match 7.2%; Score 183.5; DB 1; Length 1680;
Best Local Similarity 19.3%; Pred. No. 1.2e-05;
Matches 99; Conservative 49; Mismatches 152; Indels 213; Gaps 26;
QY 23 CPVETETWTAQVDDIGTPANCVCOKNFY-----NN 55
| ||| : ||
DB 1051 CPDGYFENSRTCTVCEP-NCASCQDPEYCTSCDHLVHHEHKYCYSACPLDTYETDN 1109
| ||| : ||
QY 56 AAAPVPGASTCTCPQKKDAGAQPNPATANLVT-----QCNVKCPAGTATAGGA 105
| ||| : ||
DB 1110 KCAFC--HSTCATC-----NGPTDQDCITCRSSRYAWQNKCLISCPDGF----- 1151
| ||| : ||
QY 106 TDYA-AITTECVNCRINFYNNAPNFNAGASTCTA-----CPVNRVG 146
| ||| : ||
DB 1152 --YADKKRLECMPCQ-----EGCKTCTSNVCSECLQNWLNRKDKCIYSGSE 1197
| ||| : ||
QY 147 GALTAGNAATIVAQCN-----VACPTGTALDDGVTFDYVRSFTECVK- 188
| ||| : ||
DB 1198 GCSESEFFSVQEGQCRPCHASGCSNGPADTSCISCPNRLLEQ-----SRCVSG 1247
| ||| : ||
QY 189 CRLNFIYNGNNTFPNPGKSCQTPC-----PAIKPANVAQATLGNDAITACQNVACPD 243
| ||| : ||
DB 1248 CREGPFVEAG-----SLCSPLHTCSQCVSRITCNSCKGLELQ-NGECRTTCAD 1296
| ||| : ||
QY 244 GTISAAGVNNWAQNTCTCTNCAPNFYNNAPNFPNGNITCLPCPANKDYGAEATAGGAAT 303
| ||| : ||
DB 1297 GYSDRGI-----CAKCYLSCHTCSGPRN-----QCVCQCPA-----GWOLAAAG----- 1335
| ||| : ||
QY 304 LAKQCNIACPDGTATASGATNYVILQTECLNCAANFYDGNNGFQAGSRCKRCAANKVQG 363
| ||| : ||
DB 1336 ---ECHPECEG-----FYKSDFGCQKC--HHYCKTCN-DAGPLACTSCPPHSM-- 1378
| ||| : ||
QY 364 AVATAGGTATLIAQCALECPACTVLTDTGTSTYK----- 397
| ||| : ||
```


Db 1379 -----LDGGLCMC-LSSQYDTSATKCHDSCRSFGPGQFSCKGCVPLHL 1427

QY 398 -QAASECVCRAANFTTQTOWAGIDTCTSCN 429

Db 1428 DQNSQCVSCCQNTLAETSSAA-----CCNCD 1456

RESULT 9

PK5_MOUSE

AC Q04592; Q62040; STANDARD; PRT; 1877 AA.

DT 01-FEB-1995 (Rel. 31, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE (Protein convertase PC5) (Subtilisin/kexin type 5 precursor (PC 3.4.21.-))

DE (Convertase PC5) (PC6) (Subtilisin-like protein convertase 6) (SPC6).

DE (SPC6).

GN PCSK5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).

RC STRAIN=ICR; TISSUE=Intestine;

RX MEDLINE=93327934; PubMed=8335106;

RA Nakagawa T., Murakami K., Nakayama K.;

RT "Identification of an isoform with an extremely large Cys-rich region of PC6, a Kex2-like processing endoprotease.";

RL FEBS Lett. 327:165-171(1993).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM PC5A).

RC TISSUE=Brain, and Intestine;

RX MEDLINE=93224489; PubMed=8468318;

RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K., Nakayama K.;

RT "Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACE4.";

RL J. Biochem. 113:132-135(1993).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM PC5A).

RC TISSUE=adrenal cortex;

RX MEDLINE=93342056; PubMed=8341687;

RA Lussan J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;

RT "CDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).

RN [4]

RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.

RX MEDLINE=97103178; PubMed=8947550;

RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K., Bendayan M., Seidah N.G.;

RT "The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments.";

RL J. Cell Biol. 135:1261-1275(1996).

RN [5]

RP DEVELOPMENTAL EXPRESSION.

RX MEDLINE=96293359; PubMed=8698813;

RA Constam D.B., Calton M., Robertson E.J.;

RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";

RL J. Cell Biol. 134:181-191(1996).

RN [6]

RP DEVELOPMENTAL EXPRESSION.

RX MEDLINE=97436919; PubMed=9291583;

RA Rancourt S.L., Rancourt D.E.;

RT "Murine subtilisin-like proteinase SPC6 is expressed during embryonic implantation, somitogenesis, and skeletal formation.";

RL Dev. Genet. 21:75-81(1997).

RN [7]

RP FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY

CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION OF GROWTH FACTORS.

CC -|- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.

CC -|- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH EARLY ENDOSOMES.

CC -|- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PC5B/LONG (SHOWN HERE) AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -|- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.

CC -|- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO, EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE RIBS AND THE LIVER, BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK SAC FOLLOWED BY A CONFINATION TO DERMATOTOME COMPARTMENT. BETWEEN E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED TO THE CONDENSING MESENCHYM SURROUNDING THE CARTRIDGE. AT THIS STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL CARTRIDGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5, ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI.

CC ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF ISOFORM B OCCUR AT E12.5.

CC -|- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC RETICULUM.

CC -|- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS WITH THE TGN SORTING PROTEIN PACS-1.

CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

CC -|- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D17583; BAA04507.1; -

CC EMBL; D12619; BAA02143.1; -

CC EMBL; L14932; AAY4636.1; -

CC PIR; JX0248; JX0248.

CC PIR; A48225; A48225.

CC HSSP; Q99405; 1MPT.

CC MEROPS; S08.076; -

CC MGD; MGI:97515; Pcsk5.

CC InterPro; IPR000561; EGF-like.

CC InterPro; IPR002174; Furin-like.

CC InterPro; IPR002884; P_domain.

CC InterPro; IPR00209; Peptidase_S8.

CC Pfam; PF00082; Peptidase_S8; 1.

CC Pfam; PF01483; P_PARTIAL.

CC PRINTS; PR00723; SUBTILISIN.

CC ProDom; PD000717; P_domain; 1.

CC SMART; SM00181; EGF; 3.

CC SMART; SM00001; EGF-like; 2.

CC SMART; SM00261; FU_22

CC PROSITE; PS00136; SUBTILASE_ASP; 1.

CC PROSITE; PS00137; SUBTILASE_HIS; 1.

CC PROSITE; PS00138; SUBTILASE_SER; 1.

CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;

CC Cleavage on pair of basic residues; Repeat; Alternative splicing;

KW

[illegible]

```

Db 1132 AVTTVGTCPYVTGTGLTDLICAAAY-----NANCTANKACTACQEKKATCNLY 1179
QY 212 ---TPCAIKPANVAQATLGNDAITACQNVACPDGIIISAAGVNNVY--AQNTECTNCAP 266
Ra Baumbgartner R., Cziisch M., Mayer U., Poeschl E., Huber R.,
Db 1180 TTEATCSTSAANATADKCAWSGAACLAVTTVTATECAVYTGTLTDLICAAYNANCT--- 1235
Rt Timpl R., Holak T.A.;
Rt "Structure of the nidogen binding LE module of the laminin gammal
chain in solution.";
J. Mol. Biol. 257:658-668(1996).
QY 267 NFYNNAPFNPNSTCLPCPANKD--YCAEAT--AGGAATLAKQC---NIACPDGTAIA- 319
Db 1236 -----ANKACTACQEKKATCNLYTTEATCSTSAANATADKCAWSGAACLAVTTVAT 1286
QY 320 -----SGATNYVILQTECLNCAANFYFDGNNFQAGSSRCKACAPANK----- 360
Db 1287 ECAYVTGTGLTN-AICAAYNANCTAN-----KAGT-----ACQEKKATCNLYTTEATC 1333
QY 361 -VOGAVATA-----GGTATLIAQCALECP--AGTVLTGDTTSTY-----KQAASEC-- 403
Db 1334 STSAANATADKCAWSGAACLAVTTVATECAVYTGTLTNAICAAYNANCTANKAGTACQE 1393
QY 404 VKAANFYTKQTDWVAGIDTCTSCNKKLTSGAEANLPESAKKNIOQ 450
Db 1394 KKATCNLYTTEAT-----CS---TSAANATADKCAWSGAAC 1426

RESULT 11
LMGL_MOUSE STANDARD; PRT; 1607 AA.
AC P02468;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin gamma-1 chain precursor (Laminin B2 chain).
GN LAMC1 OR LAMB-1 OR LAMB-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88059118; PubMed=3680290;
RA Sasaki M., Yamada Y.;
RT "The laminin B2 chain has a multidomain structure homologous to the
B1 chain.";
RL J. Biol. Chem. 262:17111-17117(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89000737; PubMed=3167041;
RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;
RT "Primary structure of the mouse laminin B2 chain and comparison with
laminin B1.";
RL Biochemistry 27:5198-5204(1988).
RN [3]
RP SEQUENCE OF 1-239 FROM N.A.
RX MEDLINE=88228071; PubMed=2836421;
RA Ogawa K., Burdello P.D., Sasaki M., Yamada Y.;
RT "The laminin B2 chain promoter contains unique repeat sequences and
is active in transient transfection.";
RL J. Biol. Chem. 263:8384-8389(1988).
RN [4]
RP SEQUENCE OF 1391-1607 FROM N.A.
RX MEDLINE=85051302; PubMed=6209134;
RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
coiled-coil alpha-helix.";
RL EMBO J. 3:2355-2362(1984).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.
RX MEDLINE=96196434; PubMed=8648630;
RA Stetefeld J., Mayer U., Timpl R., Huber R.;
RT "Crystal structure of three consecutive laminin-type epidermal growth
factor-like (LE) modules of laminin gammal chain harboring the
nidogen binding site.";
RL J. Mol. Biol. 257:644-657(1996).

```

```

RN [6]
RP STRUCTURE BY NMR OF 824-881.
RX MEDLINE=96196435; PubMed=8648631;
Ra Baumbgartner R., Cziisch M., Mayer U., Poeschl E., Huber R.,
Rt Timpl R., Holak T.A.;
Rt "Structure of the nidogen binding LE module of the laminin gammal
chain in solution.";
J. Mol. Biol. 257:658-668(1996).
QY 1- FUNCTION: Binding to cells via a high affinity receptor, laminin
is thought to mediate the attachment, migration, and organization
of cells into tissues during embryonic development by interacting
with other extracellular matrix components.
QY 2- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
different polypeptide chains (alpha, beta, gamma), which are bound
to each other by disulfide bonds into a cross-shaped molecule
comprising one long and three short arms with globules at each
end.
QY THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),
LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).
QY 1- SUBCELLULAR LOCATION: Extracellular.
QY 1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
COMPONENT).
QY 1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
QY 1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
QY 1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
QY 1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.
QY 1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
QY This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
QY -----
QY EMBL; X05211; CAA28838.1; -
QY EMBL; J03484; AAA39405.1; -
QY EMBL; J02930; AAA39408.1; -
QY EMBL; J03749; AAA39409.1; -
QY PIR; A28469; MMSB2.
QY PDB; 1KLO; 20-AUG-97.
QY PDB; 1TLE; 12-FEB-97.
QY MGD; MGI:99914; Lamc1.
QY InterPro: IPR004089; Chntaxis_transd.
QY InterPro: IPR000561; EGF-like.
QY InterPro: IPR001886; LamNT.
QY InterPro: IPR000034; Laminin_B.
QY InterPro: IPR002049; Laminin_EGF.
QY Pfam; PF00052; laminin_B; 1.
QY Pfam; PF00053; laminin_EGF; 10.
QY Pfam; PF00055; laminin_Nterm; 1.
QY PRINTS; PR00011; EGFLAMININ.
QY ProDom; PD002082; LamNT; 1.
QY ProDom; PD003031; Laminin_B; 1.
QY SMART; SM00180; EGF_Lam; 9.
QY SMART; SM00001; EGF_Like; 1.
QY SMART; SM00281; Lamb; 1.
QY SMART; SM00136; LamNT; 1.
QY PROSITE; PS00022; EGF_1; 8.
QY PROSITE; PS01186; EGF_2; 2.
QY PROSITE; PS01248; LAMININ_TYPE_EGF; 10.
QY Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
QY SIGNAL 1 33
QY CHAIN 34 1607 LAMININ GAMMA-1 CHAIN.
QY DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).
QY DOMAIN 284 339 LAMININ EGF-LIKE 1.
QY DOMAIN 340 395 LAMININ EGF-LIKE 2.
QY DOMAIN 396 442 LAMININ EGF-LIKE 3.
QY DOMAIN 443 492 LAMININ EGF-LIKE 4.

```

FT	DOMAIN	493	502	LAMININ EGF-LIKE 5 (N-TERMINAL).	FT	CONFLICT	1475	1475	R -> K (IN REF. 4).
FT	DOMAIN	503	687	LAMININ DOMAIN IV.	FT	CONFLICT	1576	1576	D -> N (IN REF. 4).
FT	DOMAIN	688	721	LAMININ EGF-LIKE 5 (C-TERMINAL).	SQ	SEQUENCE	1607	AA; 177297	MW; 81B7B08E4869F242 CRC64;
FT	DOMAIN	722	770	LAMININ EGF-LIKE 6.	Query Match				
FT	DOMAIN	771	825	LAMININ EGF-LIKE 7.	Best Local Similarity 21.5%; Pred. No. 2.6e-05;				
FT	DOMAIN	826	881	LAMININ EGF-LIKE 8 (NIDOGEN-BINDING).	Matches 103; Conservative 45; Mismatches 175; Indels 155; Gaps 29;				
FT	DOMAIN	882	932	LAMININ EGF-LIKE 9.	Query Match				
FT	DOMAIN	933	980	LAMININ EGF-LIKE 10.	Score 178.5; DB 1; Length 1607;				
FT	DOMAIN	981	1028	LAMININ EGF-LIKE 11.	7.0%;				
FT	DOMAIN	1029	1607	DOMAIN II AND I.	15 INOIKSANCPCVGTETNTAGVDDL-----GTPANCV-NCQKNFYNNAAAFVPGAS 64				
FT	DOMAIN	1034	1594	COILED COIL (POTENTIAL).	646 LNNLTISIKIRGISYERTAGYLDVTLQSSARPGVGPATVWVESCPCVGYGGQF-----CE 700				
FT	DISULFID	340	349	BY SIMILARITY.	65 TCTPCPOKKDAGAAPNPATANLVTCN---VKCPAGTAIAGGATDYAAIITECVNCRIN 121				
FT	DISULFID	342	365	BY SIMILARITY.	701 TCLPGYRRETPLSLGPSPC---VLCTCNHSETCDPTGVCDGCRDNTAG--PHCEKCSDG 755				
FT	DISULFID	368	377	BY SIMILARITY.	122 FYNENAPNENAGASTCTACPVNRVGGALTAGNAATIVAQC-----VACPTGTA-----L 171				
FT	DISULFID	380	393	BY SIMILARITY.	756 YYGDSI---LGTSSDCQPCP-----CPGGSSCAIVPKTKVVCVTHCPTGTAGKRCELC 805				
FT	DISULFID	396	408	BY SIMILARITY.	172 DGGVTDYVRSFTCEVKCRLNFYNNNGNNTFPNPKSOCTPCPAIKPANVAQATLGND 231				
FT	DISULFID	414	425	BY SIMILARITY.	806 DCGYFGDPL-----GSGFVRL-----CRPCQC---NDNIDPNAVGN 841				
FT	DISULFID	428	440	BY SIMILARITY.	232 TITAOQ-----NVA-----CPDGTISAGVNNWNAON---TECTNCAPNFYN--NNAPNF 276				
FT	DISULFID	443	454	BY SIMILARITY.	842 RLTGECCLKIYNTAGFYCDRCKEGFF-----GNPLAPNADKCKACACPNYGTGVQOQSSC 896				
FT	DISULFID	445	461	BY SIMILARITY.	277 NP--GNSSTCLPCPANKD-----YGAETAG-----GAATLAKQCN-----ACPD 314				
FT	DISULFID	463	472	BY SIMILARITY.	897 NPVTGQCQCLPHVSGRDCGTPGYYNLQSGQGCRCDCGHALGSTNGQCDIRTGQCECOP 956				
FT	DISULFID	475	490	BY SIMILARITY.	315 GTAIASGATNVILOTECLNCAANFYFDGNNFQAGSSRCCKACPAKPVQGAATAGGTATL 374				
FT	DISULFID	722	731	BY SIMILARITY.	957 G-----ITGQHCERC-----ETNHFGEFGEGCKPCDCHH-----EGSL 989				
FT	DISULFID	724	738	BY SIMILARITY.	375 IAOQ-----ALECPAGTVLTDGTTSTYKQAAASECVKCAANFYTTKQTDWVAGIDTFTSC 428				
FT	DISULFID	740	749	BY SIMILARITY.	990 SLQCKDGRCECREGFV-----GNRCDQCEENFYNNR--SW-PGQCECPAC 1032				
FT	DISULFID	752	768	BY SIMILARITY.	RESULT 12				
FT	DISULFID	771	779	BY SIMILARITY.	LMAL_HUMAN				
FT	DISULFID	773	790	BY SIMILARITY.	ID LMAL_HUMAN STANDARD; PRT; 3075 AA.				
FT	DISULFID	793	802	BY SIMILARITY.	AC P25391;				
FT	DISULFID	805	823	BY SIMILARITY.	DT 01-MAY-1992 (Rel. 22, Created)				
FT	DISULFID	826	840	BY SIMILARITY.	DT 01-MAY-1992 (Rel. 22, Last sequence update)				
FT	DISULFID	828	847	BY SIMILARITY.	DT 15-JUN-2002 (Rel. 41, Last annotation update)				
FT	DISULFID	850	859	BY SIMILARITY.	DE Laminin alpha-1 chain precursor (Laminin A chain).				
FT	DISULFID	862	879	BY SIMILARITY.	GN LAM1 OR LAMA.				
FT	DISULFID	882	896	INTERCHAIN (PROBABLE).	OS Homo sapiens (Human).				
FT	DISULFID	884	903	INTERCHAIN (PROBABLE).	OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
FT	DISULFID	905	914	INTERCHAIN (WITH CHAIN BETA-1).	OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
FT	DISULFID	917	930	N-LINKED (GLCNAC. . .) (POTENTIAL).	OX NCBI_TaxID=9606;				
FT	DISULFID	933	945	N-LINKED (GLCNAC. . .) (POTENTIAL).	RN [1]				
FT	DISULFID	935	952	N-LINKED (GLCNAC. . .) (POTENTIAL).	RP SEQUENCE FROM N.A.				
FT	DISULFID	954	963	N-LINKED (GLCNAC. . .) (POTENTIAL).	RX MEDLINE=91333420; PubMed=1714537;				
FT	DISULFID	966	978	N-LINKED (GLCNAC. . .) (POTENTIAL).	RA Haaparanta T., Uitto J., Ruoslahti E., Engvall E.;				
FT	DISULFID	981	993	N-LINKED (GLCNAC. . .) (POTENTIAL).	RT "Molecular cloning of the cDNA encoding human laminin A chain.";				
FT	DISULFID	983	999	N-LINKED (GLCNAC. . .) (POTENTIAL).	RL Matrix 11:151-160(1991).				
FT	DISULFID	1001	1010	N-LINKED (GLCNAC. . .) (POTENTIAL).	RN [2]				
FT	DISULFID	1013	1026	N-LINKED (GLCNAC. . .) (POTENTIAL).	RP SEQUENCE OF 1-2628 FROM N.A.				
FT	DISULFID	1029	1029	N-LINKED (GLCNAC. . .) (POTENTIAL).	RX MEDLINE=91264789; PubMed=2049067;				
FT	DISULFID	1032	1032	N-LINKED (GLCNAC. . .) (POTENTIAL).	RA Nissinen M., Vuolteenaho R., Boot-Handford R., Kallunki P.,				
FT	DISULFID	1598	1598	N-LINKED (GLCNAC. . .) (POTENTIAL).	RA Tryggvason K.;				
FT	CARBOHYD	58	58	G -> A (IN REF. 3).	RT "Primary structure of the human laminin A chain. Limited expression				
FT	CARBOHYD	132	132	E -> D (IN REF. 2).	FT in human tissues.";				
FT	CARBOHYD	574	574	S -> C (IN REF. 2).	RL Biochem. J. 276:369-379(1991).				
FT	CARBOHYD	648	648	LR -> PS (IN REF. 2).	RN [3]				
FT	CARBOHYD	1020	1020	D -> Y (IN REF. 2).	RP SEQUENCE OF 2397-3072 FROM N.A.				
FT	CARBOHYD	1105	1105	T -> S (IN REF. 2).	RX MEDLINE=89280632; PubMed=2733383;				
FT	CARBOHYD	1159	1159	MISSING (IN REF. 2).	RA Olsén D., Nagayoshi T., Fazio M.,				
FT	CARBOHYD	1173	1173	V -> A (IN REF. 2).	RA Sanborn D., Sasaki T., Kuivaniemi H., Chu M.L., Deutzmann R.,				
FT	CARBOHYD	1203	1203						
FT	CARBOHYD	1221	1221						
FT	CARBOHYD	1239	1239						
FT	CARBOHYD	1378	1378						
FT	CARBOHYD	1393	1393						
FT	CARBOHYD	1437	1437						
FT	CONFLICT	216	216						
FT	CONFLICT	260	260						
FT	CONFLICT	337	337						
FT	CONFLICT	447	448						
FT	CONFLICT	544	544						
FT	CONFLICT	662	662						
FT	CONFLICT	886	886						
FT	CONFLICT	1158	1158						
FT	CONFLICT	1434	1434						


```
FT DISULFID 1452 1466 BY SIMILARITY.
FT DISULFID 1454 1476 BY SIMILARITY.
FT DISULFID 1479 1488 BY SIMILARITY.
FT DISULFID 1491 1506 BY SIMILARITY.
FT DISULFID 1509 1521 BY SIMILARITY.
FT DISULFID 1511 1528 BY SIMILARITY.
FT DISULFID 1530 1539 BY SIMILARITY.
FT DISULFID 1542 1553 INTERCHAIN (PROBABLE).
FT DISULFID 1556 1560 INTERCHAIN (PROBABLE).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 555 555 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 763 763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1407 1407 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1579 1579 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1596 1596 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1678 1678 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1689 1689 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1698 1698 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1717 1717 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.9%; Score 175.5; DB 1; Length 3075;
Best Local Similarity 21.2%; Pred. No. 8.1e-05;
Matches 110; Conservative 44; Mismatches 187; Indels 177; Gaps 30;

QY 18 IKSANCPVGTENTAGVDLDTGPANVCNCKNFYNNAAAFVPGASTCTPCPKKDKAGA 77
Db 706 VEHCEPQGY-----TGTSCSCLSGYRVGDILF---GGICQPCCHGHA-- 748
QY 78 QNPPTANLVTCQNVKCPAGTAAGATDYAAIITECVNCGRINFRYNNAPNENAGATC 137
Db 749 -----AECNVH---GVCI---ACAHNTTGVHCEQCLPGYGPSPRGTFGDCQPC 791
QY 138 TACPVRVGGALTAGNAATIAQCNVACPTGTALDDG--VTVDYVR---SFTCECVKRLN 192
Db 792 -ACPL-----TIAS--NFSPT-CHLNDGDEVVCDWCAPGYSANWCERCADG 834
QY 193 FYNNGNNTFPNPKSOCTPCPAIKPANVAQATLGNDAITTAQCNVACPDGTTISAAGVN 252
Db 835 YYGN-----FTVPGES-CVPCDC--SGNVDPSEAGHCDSVTGEC-LKCLGNTDGA---- 880
QY 253 NWVAQTECTNCAPFYNN--NAPN-----FNGNSTCLPCPANKD-- 291
Db 881 -----HCERCADGFYGDVATNMCRCACECHVKGSHSAVCHLETGLCDCKPNTVGTQOCD 933
QY 292 -----YGAETAGGAATLAKQCNIACPDGTATIASGATNY-----VILQTECLNCAAN 338
Db 934 QCLHGYGLDSHG-----CRPCN--CSVAGSVSDGCTDEGQCHVCPVAGKRCDCRCAHG 986
QY 339 FFDNNFQAQSSRCKAPANKVQAVATAGTATLIAQCALECPAGT-----VLTDGT 392
Db 987 FY-----AYQDGS--CTPCDCPHTQNTCDPETG-----ECV--CPPHQTGGKCECEDGH 1032
QY 393 TSTYKQAASECVKCAANYTTKTQDWAGIDTCTG-----AKKNI---QCD 451
Db 1033 WYDEVCQACNCSLGVSTHRRCDVVTHGQCKSKFGGRACDQCSGLGYRDFPCVPCDC 1092
QY 429 NKLTSGAEALPES-----AKKNI---QCD 451
Db 1093 DLRGTSGDACNLEQGLCGCVETGACPKCNVFGPQCN 1130

RESULT 13
ID LMA5_MOUSE STANDARD; PRT; 3718 AA.
AC Q61001; 09JH06;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DE Laminin alpha-5 chain precursor.
GN LAMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-92 FROM N.A.
RA Timpl R., Sasaki T.;
RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
chain.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[2]
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RC MEDLINE=96081906; PubMed=7499364;
RX Miner J.H., Lewis R.M., Sanes J.R.;
RA "Molecular cloning of a novel laminin chain, alpha 5, and widespread
expression in adult mouse tissues.";
RL J. Biol. Chem. 270:28523-28526(1995).
[3]
RP REVISIONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RC Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
is thought to mediate the attachment, migration, and organization
of cells into tissues during embryonic development by interacting
with other extracellular matrix components.
CC -1- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
different polypeptide chains (alpha, beta, gamma), which are bound
to each other by disulfide bonds into a cross-shaped molecule
Comprising one long and three short arms with globules at each
end.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
MEMBRANES (MAJOR COMPONENT).
CC -1- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ293593; CAB99255.1; -.
DR HSSP; U37501; AAC53430.1; -.
DR HSSP; P02468; ITLE.
DR MGD; MGI:105382; Lama5.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00052; laminin_B; 1.
DR Pfam; PF00053; laminin_EGF; 19.
DR Pfam; PF00054; laminin_G; 2.
DR Pfam; PF00055; laminin_Nterm; 1.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 1.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS01186; EGF_2; 3.
```

DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
DR PROSITE; ESS0025; LAM_G_DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.
FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 305 363 LAMININ EGF-LIKE 1.
FT DOMAIN 364 433 LAMININ EGF-LIKE 2.
FT DOMAIN 434 479 LAMININ EGF-LIKE 3.
FT DOMAIN 500 546 LAMININ EGF-LIKE 4.
FT DOMAIN 547 592 LAMININ EGF-LIKE 5.
FT DOMAIN 593 637 LAMININ EGF-LIKE 6.
FT DOMAIN 638 682 LAMININ EGF-LIKE 7.
FT DOMAIN 683 728 LAMININ EGF-LIKE 8.
FT DOMAIN 729 781 LAMININ EGF-LIKE 9.
FT DOMAIN 782 833 LAMININ EGF-LIKE 10.
FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.
FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
FT DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
FT DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
FT DOMAIN 1633 1642 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1643 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
FT DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
FT DOMAIN 2169 2735 DOMAIN II AND I.
FT DOMAIN 2736 2933 LAMININ G-LIKE 1.
FT DOMAIN 2947 3119 LAMININ G-LIKE 2.
FT DOMAIN 3128 3296 LAMININ G-LIKE 3.
FT DOMAIN 3337 3511 LAMININ G-LIKE 4.
FT DOMAIN 3518 3689 LAMININ G-LIKE 5.
FT DOMAIN 2205 2257 COILED COIL (POTENTIAL).
FT DOMAIN 2330 2464 COILED COIL (POTENTIAL).
FT DOMAIN 2604 2621 COILED COIL (POTENTIAL).
FT DOMAIN 2639 2705 COILED COIL (POTENTIAL).
FT SITE 1723 1725 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1839 1841 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 305 314 BY SIMILARITY.
FT DISULFID 307 327 BY SIMILARITY.
FT DISULFID 329 338 BY SIMILARITY.
FT DISULFID 341 361 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 366 398 BY SIMILARITY.
FT DISULFID 401 410 BY SIMILARITY.
FT DISULFID 413 431 BY SIMILARITY.
FT DISULFID 434 445 BY SIMILARITY.
FT DISULFID 436 452 BY SIMILARITY.
FT DISULFID 454 463 BY SIMILARITY.
FT DISULFID 466 476 BY SIMILARITY.
FT DISULFID 500 512 BY SIMILARITY.
FT DISULFID 502 521 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 535 544 BY SIMILARITY.
FT DISULFID 547 559 BY SIMILARITY.
FT DISULFID 549 566 BY SIMILARITY.
FT DISULFID 568 577 BY SIMILARITY.
FT DISULFID 580 590 BY SIMILARITY.
FT DISULFID 593 605 BY SIMILARITY.
FT DISULFID 595 611 BY SIMILARITY.
FT DISULFID 613 622 BY SIMILARITY.
FT DISULFID 625 635 BY SIMILARITY.
FT DISULFID 638 650 BY SIMILARITY.
FT DISULFID 640 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 670 680 BY SIMILARITY.
FT DISULFID 683 695 BY SIMILARITY.
FT DISULFID 702 713 BY SIMILARITY.
FT DISULFID 716 726 BY SIMILARITY.
FT DISULFID 1443 1455 BY SIMILARITY.
FT DISULFID 1445 1462 BY SIMILARITY.
FT DISULFID 1464 1473 BY SIMILARITY.
FT DISULFID 1476 1486 BY SIMILARITY.
FT DISULFID 1533 1548 BY SIMILARITY.
FT DISULFID 1555 1566 BY SIMILARITY.
FT DISULFID 1557 1579 BY SIMILARITY.
FT DISULFID 1569 1579 BY SIMILARITY.
FT DISULFID 1582 1594 BY SIMILARITY.
FT DISULFID 1584 1601 BY SIMILARITY.
FT DISULFID 1603 1612 BY SIMILARITY.
FT DISULFID 1615 1630 BY SIMILARITY.
FT DISULFID 1665 1674 BY SIMILARITY.
FT DISULFID 1867 1881 BY SIMILARITY.
FT DISULFID 1884 1893 BY SIMILARITY.
FT DISULFID 1896 1912 BY SIMILARITY.
FT DISULFID 1915 1930 BY SIMILARITY.
FT DISULFID 1917 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1953 1968 BY SIMILARITY.
FT DISULFID 1971 1986 BY SIMILARITY.
FT DISULFID 1973 1993 BY SIMILARITY.
FT DISULFID 1996 2005 BY SIMILARITY.
FT DISULFID 2008 2022 BY SIMILARITY.
FT DISULFID 2072 2083 BY SIMILARITY.
FT DISULFID 2074 2090 BY SIMILARITY.
FT DISULFID 2092 2101 BY SIMILARITY.
FT DISULFID 2104 2116 BY SIMILARITY.
FT DISULFID 2119 2126 BY SIMILARITY.
FT DISULFID 2121 2133 BY SIMILARITY.
FT DISULFID 2135 2144 BY SIMILARITY.
FT DISULFID 2147 2166 BY SIMILARITY.
FT DISULFID 2169 2199 INTERCHAIN (PROBABLE).
FT DISULFID 2172 2172 INTERCHAIN (PROBABLE).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 905 905 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1335 1335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2021 2021 N-LINKED (GLCNAC. . .) (POTENTIAL).
Query Match 6.9%; Score 174.5; DB 1; Length 3718;
Best local Similarity 20.8%; Pred. No. 0.00012;
Matches 93; Conservative 34; Mismatches 140; Indels 181; Gaps 25;
Qy 32 AGQVDDLGTAN-----CVNCQKNFYNNAAAFVPGASTCTPCPKQKADAGQNPATAN 86
Db 1826 ASNVELCMCPANYRGDSQCECAPGYRDTKGLFL---GRCVPC----- 1865
Qy 87 LVTOC---NVKCPAGTAAGGATDYAAITTECVNCRINFYNNAPNNAGASTCTACPVN 143
Db 1866 ---QCHGHSRDLCPGSGICVGC-QHNTEGQCRCRCPGVSSDPN---PASPCVSCP-- 1916
Qy 144 RVGGALTAGNATVAOCNVACPT-----GTALDDGVTYDVR---SFTCEVCRLNFY 195
Db 1917 -----CPLAVSNFNFAADGCVLRNGRTQCLCRPGYAGASCERCAPGFFG 1959
Qy 196 NGNNGTFFNPGKSOCTPCPAIKPANVAQATLGNDATTITACNV---ACPDGTISAAGVN 252
Db 1960 N-----PLVLG-SSCQPCDC-----SGNGDPNNIFSDCDPLTGACGCLRHITG-- 2002
Qy 253 NWVAQNTCTNCAPNFYNNAPNPNPNST---CLPCPANKDYGAETAGATLAKQC- 308
Db 2003 -----PHCERCAPGFYGN---ALLPGNCTRCDCSPC-----GTETCDPSG 2040

```

QY 309 NIACPDGTAIASGATNYVILQTECLNCAANFYFDGNFQAGSSRCKAC-----PANKVQ 362
DQ 2041 RCLCRAG-----VTGQRCDRCLEGYF-----GFEOCQGRPCACGPAARGS 2081
QY 363 GAVATAGGTATLIAQCALECPAGVLTDTGTYSTYKQARASECVKCAANFYTKQTDWVAGI 422
DQ 2082 ECHPOSQ-----QC--HCQPGT-----TGPOCLECAPGTW----- 2109
QY 423 DTCTSCNKKLTSGABANLPESAKNNIQ 450
DQ 2110 -----GLPEKGCRCQC 2121

RESULT 14
LMA_DROME STANDARD; PRT; 3712 AA.
AC 000174;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha chain precursor.
GN LANA OR LAMA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93049203; PubMed=1425586;
RA Kusche-Gullberg M., Garrison K., Mackrell A.J., Fessler L.I.,
RA Fessler J.H.;
RT "Laminin A chain: expression during Drosophila development and
RT genomic sequence."
RL EMBO J. 11:4519-4527(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94038678; PubMed=8223265;
RA Henchcliffe C., Garcia-Alonso L., Tang J., Goodman C.S.;
RT "Genetic analysis of laminin A reveals diverse functions during
RT morphogenesis in Drosophila."
RL Development 118:325-337(1993).
RN [3]
RP SEQUENCE OF 1762-3712 FROM N.A.
RX MEDLINE=92078147; PubMed=1744083;
RA Garrison K., Mackrell A.J., Fessler J.H.;
RT "Drosophila laminin A chain sequence, interspecies comparison, and
RT domain structure of a major carboxyl portion."
RL J. Biol. Chem. 266:22899-22904(1991).
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -!- FUNCTION: DIVERSE FUNCTIONS DURING MORPHOGENESIS IN DROSOPHILA.
CC COMPLETE LOSS-OF-FUNCTION MUTATIONS LEAD TO LATE EMBRYONIC
CC LETHALITY. CERTAIN PARTIAL LOSS-OF-FUNCTION MUTATIONS GIVE RAISE
CC TO ESCAPER ADULTS, WHICH HAVE ROUGH EYES ASSOCIATED WITH CHANGES
CC IN CELL FATE AND PATTERN, MISSHAPEN LEGS AND DEFECTS IN WING
CC STRUCTURE.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: NEWLY FORMED MESODERM AND LATER PROMINENTLY
CC EXPRESSED IN HEMOCYTES, WHICH ALSO SYNTHESIZE COLLAGEN IV.
CC -!- DEVELOPMENTAL STAGE: DURING MORPHOGENESIS, MOSTLY IN EMBRYO
CC DEVELOPMENT AT 10-12 HOURS.
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT

```

```

CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV' IS NOT
CC SIMILAR TO LAMININ DOMAIN IV).
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M96388; AAA28662.1; -
CC EMBL; L07288; AAC37178.1; -
CC EMBL; M75882; AAA28661.1; -
CC HSP; P02468; 1TLE.
CC FlyBase; FBgn0002526; LANA.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR000034; Laminin_B.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR001230; Prenyl_site.
CC Pfam; PF00052; laminin_B_1.
CC Pfam; PF00053; laminin_EGF; 20.
CC Pfam; PF00054; laminin_G; 5.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRINTS; PR00011; EGFLAMININ.
CC ProDom; PD002082; LamNT; 1.
CC ProDom; PD003031; Laminin_B; 1.
CC SMART; SM00180; EGF_Lam; 17.
CC SMART; SM00001; EGF_Like; 1.
CC SMART; SM00281; LamB; 1.
CC SMART; SM00282; LamG; 5.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 17.
CC PROSITE; PS01186; EGF_2; 5.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
CC PROSITE; PS00025; LAM_G_DOMAIN; 5.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 22
FT CHAIN 23 3712 LAMININ ALPHA CHAIN.
FT DOMAIN 25 272 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 273 332 LAMININ EGF-LIKE 1.
FT DOMAIN 333 402 LAMININ EGF-LIKE 2.
FT DOMAIN 403 447 LAMININ EGF-LIKE 3.
FT DOMAIN 448 494 LAMININ EGF-LIKE 4.
FT DOMAIN 495 540 LAMININ EGF-LIKE 5.
FT DOMAIN 541 586 LAMININ EGF-LIKE 6.
FT DOMAIN 587 631 LAMININ EGF-LIKE 7.
FT DOMAIN 632 676 LAMININ EGF-LIKE 8.
FT DOMAIN 677 731 LAMININ EGF-LIKE 9.
FT DOMAIN 732 784 LAMININ EGF-LIKE 10.
FT DOMAIN 785 815 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 816 1374 DOMAIN IV'.
FT DOMAIN 1375 1420 LAMININ EGF-LIKE 12.
FT DOMAIN 1421 1465 LAMININ EGF-LIKE 13.
FT DOMAIN 1466 1513 LAMININ EGF-LIKE 14.
FT DOMAIN 1514 1564 LAMININ EGF-LIKE 15.
FT DOMAIN 1565 1574 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1575 1775 LAMININ DOMAIN IV (DOMAIN IV').
FT DOMAIN 1776 1808 LAMININ EGF-LIKE 16 (C-TERMINAL).
FT DOMAIN 1809 1858 LAMININ EGF-LIKE 17.
FT DOMAIN 1859 1916 LAMININ EGF-LIKE 18.
FT DOMAIN 1917 1969 LAMININ EGF-LIKE 19.
FT DOMAIN 1970 2016 LAMININ EGF-LIKE 20.
FT DOMAIN 2017 2063 LAMININ EGF-LIKE 21.
FT DOMAIN 2064 2111 LAMININ EGF-LIKE 22.

```


DR	EMBL; X52133; CAA36378.1; -.
DR	PTR; S09118; S09118.
DR	InrPro; IPR002895; Paramesium_SA.
DR	Pfam; PF01508; Paramesium_SA; 34.
KW	Signal; Repeat; Antigen; Membrane; GPI-anchor.
FT	SIGNAL 1 20
FT	POTENTIAL.
FT	CHAIN 21 2704
FT	DOMAIN 106 2560
FT	DOMAIN 1060 1424
FT	VARIABLE DOMAIN, COMPRISES 4 ALMOST IDENTICAL REPEATS.
SQ	SEQUENCE 2704 AA; 278775 MW; 40EAA0A0B18EE2119 CRC64;

Search completed: February 11, 2003, 19:47:08
Job time : 23.2571 secs